A STUDY OF ERRORS IN TESTS OF ADDING ABILITY[1]

C. L. PHELPS
Leland Stanford Junior University

In some of the recent tests of adding ability the results have been based almost entirely on the rapidity with which combinations were attempted and not much attention has been given to the errors that were made in taking the tests. Justification of this failure to make a careful study of errors is found in the claim that speed and accuracy show such a close positive correlation in these tests that checking for errors can safely be neglected. If this claim is true, the only remaining reason for making an analysis of errors is for diagnostic purposes. But this reason alone seemed sufficient to justify the writer in making an examination of a large amount of data in order to find out how often errors are made on the various combinations, what per cent of a group make them, how the combinations rank in difficulty, and what kind of mistakes are made on them.

SOURCE AND DESCRIPTION OF DATA

The material upon which this study is based was collected by Otis and Davidson[2] for their investigation of the reliability of standard scores as a measure of adding ability. It was collected from the eighth grade of eight of the larger grammar schools of San Jose, California, a city of about 50,000 population. It will be seen by reference to their study that their tests consisted of 25 arrangements of the combinations varying slightly in order from the Courtis Test No. 1, that the tests were one minute in length, that five were given each day for five days, that they were all given by the same person in exactly the same way, and that 270 children were thus examined.

[1] This study was suggested and guided by Professor Percy E. Davidson of Leland Stanford Junior University.

[2] "The Reliability of Standard Scores in Adding Ability," *Elementary School Teacher*, October, 1912.

Of this material the papers of 238 children have been checked for errors and the results studied. Of the 32 sets of papers rejected for this study 20 were five tests short of the full number and twelve were eliminated because they were not representative and would have vitiated the results if they had been included. For example, when it was found that the lower figure had been written for the total of several combinations in succession, it indicated that no attempt had been made to add and the sets of papers containing this error were left out. This type of error, with others which could not in fairness be included, resulted in the elimination of the twelve sets referred to as not representative. The 238 sets remaining, consisting of 5,950 tests, are considered entirely fair for the purposes of this study.

METHOD OF MARKING ERRORS AND TABULATING RESULTS

In checking the errors each paper was examined carefully and all mistakes marked. At first a separate marking was made of combinations which had been missed at the first stroke of the pencil and corrected by a second stroke. But this was later abandoned and only those errors were counted which had not been corrected at all.

When the errors were all marked, a system of tabulation was arranged in such a way that the errors on each combination could easily be totaled for each pupil as well as for the whole group. This was done for each school separately, and then the results were collected into one table which is here presented as Table I.

DISPLAY AND DISCUSSION OF TABLES

This table shows gross results just as they appeared when the first tabulation was completed. It is reproduced here in order to show how the errors were distributed among the eight schools and on the various combinations. On examination several things can be seen in regard to this point. One is that there is a rather erratic distribution of errors in the ones and twos; another is that there is a more regular distribution in the other combinations; a third is that the number of errors increases with the increase in size of the combinations; and a final one is that the errors on the doubles are almost as erratic in their distribution as on the ones and twos.

Table I takes no account, however, of the fact that in any arrangement of combinations for tests having a time limit certain

TABLE I

SHOWING ERRORS ON EACH COMBINATION FOR EACH SCHOOL AND FOR THE WHOLE GROUP

COMBINATIONS	SCHOOL No. 1	SCHOOL No. 2	SCHOOL No. 3	SCHOOL No. 4	SCHOOL No. 5	SCHOOL No. 6	SCHOOL No. 7	SCHOOL No. 8	TOTAL
	40 Pupils	12 Pupils	14 Pupils	23 Pupils	33 Pupils	48 Pupils	28 Pupils	40 Pupils	
0+0	2	2	0	0	19	0	0	0	23
0+1	1	1	0	2	5	2	3	1	15
0+2	1	1	1	1	6	1	2	3	16
0+3	0	2	1	2	5	6	1	0	17
0+4	2	1	0	1	6	1	3	0	14
0+5	4	1	2	2	9	4	7	2	31
0+6	1	1	0	0	1	0	0	0	3
0+7	0	3	0	1	5	4	1	0	14
0+8	5	1	2	2	5	12	4	6	37
0+9	0	0	0	0	4	4	1	0	9
1+1	1	4	1	1	2	1	5	1	16
1+2	1	4	0	8	8	5	4	8	38
1+3	4	2	1	7	6	18	8	7	53
1+4	8	8	7	24	22	43	12	30	154
1+5	10	16	8	11	6	16	7	12	86
1+6	22	15	4	16	21	16	21	17	132
1+7	21	17	22	19	12	19	16	18	144
1+8	6	6	2	4	3	9	5	2	37
1+9	9	6	3	4	17	11	9	6	65
2+2	6	1	3	1	4	4	0	3	22
2+3	8	7	1	4	3	3	5	1	32
2+4	6	11	2	10	5	8	7	10	59
2+5	19	9	5	8	20	14	9	18	102
2+6	7	2	3	3	10	3	11	3	42
2+7	8	17	8	11	19	32	22	41	158
2+8	3	7	1	5	8	11	6	11	52
2+9	7	6	0	7	10	13	15	13	71
3+3	15	4	2	15	5	12	19	15	87
3+4	5	1	2	1	7	13	8	3	40
3+5	10	8	14	11	9	12	8	6	78
3+6	12	15	7	6	4	11	12	4	71
3+7	8	30	11	11	7	36	11	6	120
3+8	23	34	9	21	30	44	59	11	231
3+9	48	46	16	19	39	45	52	39	304
4+4	1	2	0	0	3	0	1	0	7
4+5	3	3	2	3	4	9	12	7	43
4+6	2	12	2	7	6	4	4	4	41
4+7	21	52	12	8	8	44	50	37	232
4+8	30	23	7	14	11	21	19	17	142
4+9	32	56	23	19	11	19	22	10	192

TABLE I—*Continued*

Combinations	School No. 1	School No. 2	School No. 3	School No. 4	School No. 5	School No. 6	School No. 7	School No. 8	Total
	40 Pupils	12 Pupils	14 Pupils	23 Pupils	33 Pupils	48 Pupils	28 Pupils	40 Pupils	
5+5	0	0	1	1	1	0	1	0	4
5+6	17	23	5	10	15	18	22	25	135
5+7	16	35	6	11	16	13	26	11	134
5+8	78	63	26	51	24	44	58	25	369
5+9	32	41	21	47	19	30	78	30	298
6+6	2	1	1	3	1	7	1	5	21
6+7	81	11	10	13	14	24	12	21	186
6+8	14	15	0	1	6	3	18	5	62
6+9	32	98	7	35	23	50	42	33	320
7+7	2	3	2	1	5	1	2	0	16
7+8	35	11	16	6	9	18	19	23	137
7+9	71	75	5	38	61	64	40	41	395
8+8	14	11	8	20	6	22	25	12	118
8+9	14	27	3	13	31	21	20	16	145
9+9	3	0	0	3	3	3	1	5	18
Total	783	851	295	542	619	848	826	624	5,388

combinations will be attempted oftener than others because some pupils will work more rapidly than others. For this reason only approximate results can be obtained for those combinations which, in the arrangement of the tests, fall beyond the average number of attempts made by the pupils tested. Accurate results can be obtained, however, for all those combinations which fall safely within this limit. These will all receive practically an equal number of trials because of the slight variation in the order of the combinations for the different tests and because the average deviation from the central tendency with regard to the number of combinations is not high, except in one school where it amounts to 10.6 combinations. But in this case the average for the school is so far above the average for the whole group that results are not affected. In the other cases they would be affected but slightly, For combinations which do not fall within the range of average number of attempts, a certain amount of data is available and fairly accurate conclusions might be drawn concerning some of

them. But it is perhaps safer to eliminate them from this consideration. Combinations affected in this way are: 0+0; 1+1; 4+4; 4+9; 5+5, and 6+8. There are also certain combinations which are repeated in inverse order within the limits noted. These have all been credited with twice as many trials as the others.

These explanations prepare the reader for Table II, which is derived from Table I. In this is shown the percentage of error made on each combination, the basis for calculation being the actual number of times it was attempted. This table is considered as fair an index to the relative difficulty of the combinations as can be derived by this method from the data at hand.

TABLE II

SHOWING PERCENTAGE OF ERRORS BASED ON THE NUMBER OF TRIALS FOR EACH COMBINATION

Combinations	Errors in Percentage	Combinations	Errors in Percentage	Combinations	Errors in Percentage
0+0	.39	2+2	.37	4+8	2.38
0+1	.25	2+3	.54	4+9	*
0+2	.14	2+4	.50	5+5	.07
0+3	.28	2+5	.86	5+6	2.27
0+4	.24	2+6	.71	5+7	2.25
0+5	.52	2+7	1.33	5+8	3.10
0+6	.05	2+8	.88	5+9	2.50
0+7	.24	2+9	1.19	6+6	.35
0+8	.62	3+3	1.46	6+7	1.56
0+9	.15	3+4	.67	6+8	1.04
1+1	.27	3+5	1.31	6+9	2.60
1+2	.64	3+6	1.19	7+7	.14
1+3	.45	3+7	2.02	7+8	2.30
1+4	1.30	3+8	1.94	7+9	3.32
1+5	.72	3+9	2.55	8+8	1.98
1+6	2.22	4+4	.12	8+9	2.44
1+7	2.42	4+5	.72	9+9	.30
1+8	.62	4+6	.69		
1+9	1.09	4+7	1.95		

* Unsatisfactory data.

But it may be held by some that the method used here is not a fair one because it is based entirely on the *number of errors* made and does not take into account the *number of children* making the errors. It may further be held that, because of the large number of trials possible, a very few repeaters could make the entire number of mistakes credited to any combination. When such a possibility suggested itself all the papers were checked again to see

exactly what the facts were concerning this possibility. The findings in this search were tabulated by schools so as to show how many children missed each combination a single time and how many missed it two or more times. This table is not reproduced here, on account of lack of space, but two condensed ones, showing the rank of 20 of the combinations above the ones, are taken from it along with a similar table taken from Table II. In these three tables 20 combinations are ranked in difficulty in accord with (1) the number of times they were missed, (2) the number of children that missed them, and (3) the number of children that missed them two or more times. These tables are presented, for purposes of comparison, as Tables III*A*, III*B*, and III*C*, respectively.

TABLE III*A*

Rank	Combinations	Errors in Percentage of Attempts
1....	9+7	3.32
2....	8+5	3.10
3....	9+6	2.69
4....	9+3	2.55
5....	9+5	2.50
6....	9+8	2.44
7....	8+4	2.38
8....	8+7	2.30
9....	6+5	2.27
10....	7+5	2.25
11....	7+3	2.02
12....	8+8	1.98
13....	7+4	1.95
14....	8+3	1.94
15....	7+6	1.56
16....	3+3	1.46
17....	7+2	1.33
18....	5+3	1.31
19....	6+3	1.19
19....	9+2	1.19

TABLE III*B*

Rank	Combinations	No. of Children Making Errors	Percentage of Children Making Errors
1....	9+7	120	50.42
2....	9+3	109	45.80
3....	9+6	94	39.49
4....	8+5	91	38.23
5....	9+5	86	36.13
6....	7+6	79	33.19
6....	7+2	79	33.19
7....	8+3	74	31.09
8....	7+4	72	30.25
9....	8+8	71	29.83
10....	9+8	67	28.15
10....	8+7	67	28.15
11....	9+4	60	25.21
12....	5+2	59	24.79
13....	6+5	58	24.37
14....	8+4	52	21.85
14....	9+2	52	21.85
15....	7+5	48	20.17
15....	7+3	48	20.17
16....	6+3	44	18.48

TABLE III*C*

Rank	Combinations	No. of Children Making 2 or More Errors	Percentage of Children Making 2 or More Errors
1....	9+7	63	26.47
2....	9+3	59	24.79
3....	9+5	49	20.59
4....	8+5	47	19.75
5....	9+6	38	15.96
6....	7+4	37	15.55
6....	9+4	33	13.86
8....	8+3	32	13.44
9....	7+2	31	13.02
9....	8+7	31	13.02
10....	9+8	26	10.92
11....	7+6	25	10.50
11....	6+5	25	10.50
12....	8+8	24	10.08
13....	7+3	23	9.66
14....	8+4	21	8.82
14....	5+2	21	8.82
15....	7+5	20	8.40
16....	9+2	17	7.14
17....	8+6	15	6.30

From these tables it can be seen that there is a close correspondence in results, regardless of whether they are based on the number of times the combination was missed, on the number of children who missed it, or on the number who missed it two or more times.

It can also be seen that the higher combinations make the greatest showing of difficulty from all three points of consideration. The totals for the number of mistakes, the number of children who made them, and the number who made them repeatedly are all highest for the higher combinations.

The fact that the tests were repeated 25 times is significant in this connection. We may assume that any child might miss a combination once in 25 times on account of a slip of the pen or for some reason other than lack of knowledge of the combination, but it is not to be supposed that this would be done repeatedly. It may not be of the highest importance to note that 120 children out of 238 missed the combination 9+7, but it certainly is significant to know that 63 of them missed it two or more times. If an arbitrary standard were to be set up and it should be agreed that two or more errors in 25 chances on any combination should be taken to mean uncertainty concerning that combination, then it could be stated that from 6 per cent to 26 per cent of the children taking these tests were uncertain about some one or more of the twenty combinations included in the tables above. This would indicate that there are many children who make mistakes on the higher combinations because of lack of knowledge of them or uncertainty concerning them.

The errors on some of the combinations, especially on the ones and the doubles, seem to be out of proportion to the general indications of their difficulty, but this is because of the existence of certain *type errors*. Table IV shows the principal errors of this kind found in this study and indicates the explanation of the erratic showing of errors on certain of the combinations previously referred to.

One form of these errors is multiplication which is found mainly in the ones, though it is also found in certain other combinations. The 3+3 combination is a striking example of this. In three of the eight schools tested no other kind of mistake was made on it, while only 8 out of a total of 87 errors were due to any other cause. This kind of type error is not found in any other double, unless it is in the 2+2 where it makes no difference in the result.

In 8+8=18 a different form of type error is found which

accounts for nearly all of the mistakes made on the 8+8 combination. For example in School No. 7, 16 out of 28 children made this error one or more times.

TABLE IV

SHOWING PRINCIPAL TYPE ERRORS AS INDICATED BY THE NUMBER OF TIMES THE SAME MISTAKE WAS MADE IN EACH CASE

Combinations	School No. 1	School No. 2	School No. 3	School No. 4	School No. 5	School No. 6	School No. 7	School No. 8	Total
1+4=6.....	5	2	1	18	17	26	7	24	100
1+4=4.....	3	5	3	3	3	14	5	4	40
Miscellaneous....	0	1	3	3	2	3	0	2	14
1+6=6.....	19	9	4	11	11	11	17	11	93
1+6=8.....	3	4	0	3	7	3	3	2	25
Miscellaneous....	0	2	0	2	3	1	1	4	13
1+7=9.....	11	15	21	17	8	13	9	12	106
1+7=7.....	8	2	1	1	3	5	4	4	28
Miscellaneous....	2	0	0	1	1	1	3	2	10
2+7=7.....	4	5	5	6	9	14	13	31	87
2+7=8.....	1	11	3	4	2	8	6	3	38
2+7=11.....	0	0	0	0	3	4	2	6	15
Miscellaneous....	3	1	0	1	5	6	1	1	18
3+3=9.....	14	3	2	14	5	9	17	15	79
Miscellaneous....	1	1	0	1	0	3	2	0	8
3+9=11.....	34	24	13	12	25	35	41	27	211
3+9=13.....	2	18	0	4	5	4	3	2	38
3+9=9.....	10	2	3	2	8	3	5	9	42
Miscellaneous....	2	2	0	1	1	3	3	1	13
8+8=18.....	13	9	8	17	5	20	22	10	104
Miscellaneous....	1	2	0	3	1	2	3	2	14

The 9+3 combination also presents peculiar results. Theoretically, 9+3 should equal 13 as often as 11. Practically, it does not. In every school 11 is strongly the favorite error, while the total shows more than 5 to 1 in its favor.

Another point of interest is shown by Table V. It can be seen from this that there is a considerable difference in the number of errors made by the different groups, the average number per child varying from 15.60 in School No. 8 to 70.92 in School No. 2. It seems significant that there should be such a wide range, but the real importance of the table becomes apparent when it is noticed

that the school which had the *smallest number of errors* per pupil had the *slowest speed record*, and that the school which had the *largest number of errors* had the *highest record for speed.*

TABLE V

SHOWING AVERAGE NUMBER OF ERRORS PER PUPIL, AVERAGE NUMBER OF COMBINATIONS ATTEMPTED, AND RANK OF THE SCHOOLS ACCORDING TO SPEED AND ACCURACY

Schools	Average Number of Errors per Pupil	Rank Based on Accuracy	Average Number of Combinations per Pupil	Rank Based on Speed
1	19.57	4	69.77	5
2	70.92	8	76.08	1
3	21.07	5	71.78	3
4	23.56	6	70.13	4
5	18.76	3	73.75	2
6	17.67	2	62.04	7
7	29.50	7	68.50	6
8	15.60	1	59.82	8

From this table it is apparent that if the San Jose schools were ranked according to the number of combinations attempted in these tests and according to the errors made in the tests, the order of the schools would be almost reversed.

SUMMARY AND DISCUSSION OF RESULTS

In the introduction of this paper it was stated that this study was undertaken to find out how often errors are made on the combinations, what percentage of the group make them, how the combinations rank in difficulty, and what kind of mistakes are made on them. Something has been learned about all of these questions and the results have been presented in the foregoing tables and discussions. They may be summarized as follows:

1. Except in the zero combinations, where the errors were few, and in the ones where they were erratic because of the influence of type errors, *a fairly regular increase was found in the number of errors made and in the number of children making them as the combinations increased in size.* This finding indicates either that not enough attention has been given to the teaching of the higher combinations or that their inherent difficulty is greater than has been supposed. The fact in the case could be definitely determined

by experimentation, to the very great advantage of practice in teaching the combinations.

2. It was found that not only was there an increase in mistakes and the number of children making them as the combinations increased in size but that *a large percentage of the children* made mistakes on each of the larger combinations. On certain combinations from a *third* to a *half* made mistakes, while about half of this number made *two or more mistakes on the same combination*.

3. The rank in difficulty of the higher combinations is fairly well shown. For example, 9+7 was missed 3.32 per cent of the times it was attempted and 50.42 per cent of the children missed it. From this it can be seen that half of the children must have missed it, on the average, twice the percentage of the times credited to all. It can therefore be said that half the children taking these tests missed this combination about $6\frac{2}{3}$ per cent of the times they attempted it. Practically the same can be said of 8+5. A third or more missed 9+6, 9+3, and 9+5 an average of at least $7\frac{1}{2}$ per cent of their trials. And a fourth or more missed 9+8, 8+4, 8+7, 6+5, 7+5, 7+3, 8+8, and 7+4 an average of more than 7 per cent of their attempts. These results are derived from Tables III*A* and III*B* and, besides giving an idea of their relative difficulty, demonstrate rather clearly the lack of knowledge of the higher combinations.

4. It was found that *there were certain kinds of errors which were peculiar to certain combinations throughout all the schools tested.* These were called *type errors*. They were found to occur on several combinations and to account for a large part of the errors made on those combinations. Pedagogically, this finding is important, for, while the mere knowledge that a certain error is a *type* error may not suggest to the teacher the best method of dealing with it, it ought to prevent the ignoring of it as a slip of the pen or an accident.

In conclusion, the fact must be noted that some of the results found in this study are not in harmony with certain statements concerning tests of this kind that are widely known and relied upon. Reference is made here to two statements made by Courtis in his "Instructions for Making and Recording Scores" in giving

the Courtis Standard Tests. The first statement, which applies to four of eight of the Courtis Tests, is that "Ignorance is shown, not by incorrect answers, but by reduced speed." That is, he holds that speed and accuracy in such tests show a close positive correlation. In this study no such correlation was found. *The slowest group did the most accurate work and the fastest group made the most errors.* The second statement is that "Mistakes, except for the zero combinations, are very few." In this study, mistakes were found to be *fewest in the zero combinations and to gradually increase in number as the combinations increased in size*, with certain exceptions which have been noted. And, finally, the results of this study do not justify the ranking of groups by scoring their speed record without taking into account the errors made on the tests taken. On the other hand, they indicate that speed records alone may be unfair and sometimes directly opposed, as they were in some cases in this study, to the findings of a tabulation of errors.